

Education Week's blogs > Assessing the Assessments

Five Myths About International Large-Scale Assessments

By Invited Contributor Listed Below on March 23, 2014 11:55 PM | **2 Comments**

Today's guest contributors are **Laura C. Engel**, Assistant Professor of International Education and International Affairs, The George Washington University and **Michael J. Feuer**, Dean of the Graduate School of Education and Human Development; Professor of Education, The George Washington University.

With the recent release of the 2012 Programme for International Student Assessment (PISA) results, we are once again reminded about the extent to which international large-scale assessments (ILSAs) have gripped the world of education. ILSAs of course consist of a diverse set of assessments, ranging from math and science, reading, civic and citizenship education, teacher education, among others (for a complete list and history of ILSAs, see the introduction and appendix of **this special issue** edited by Engel and Williams, 2013). ILSAs offer exciting insights into complex education systems and serve as invaluable tools to compare education systems internationally. Yet, with their high profile and considerable policy impact, ILSAs are also surrounded by a number of persistent myths. With an aim to shrink the distance between the widespread beliefs and the emergent evidence of ILSAs, we explore five prevailing myths.

Myth #1: *Average achievement scores offer an accurate and comprehensive record of the overall quality and effectiveness of the U.S. education system.* The convenience of a single score to represent a system's performance has proved to be consistently appealing to policy-makers, the media, reformers, and the public. Researchers have warned against the inherent dangers in using a single average achievement score as the leading indicator of educational quality. A more accurate and useful picture comes instead from deeper explorations of statistically significant performance variances among participating systems (for a resource on ILSA data analyses, see the recent handbook on international large-scale assessment edited by **Rutkowski, von Davier & Rutkowski, 2013**). It is also beneficial to draw on multiple data sources, including from national assessments and other mixed-methods educational research.

Myth #2: *ILSAs prove that the U.S. education system is declining.* Scholars have disputed the many zealous accounts of a stagnating or declining U.S. education system. Some argue that the U.S. has never actually been first in the world educationally, pointing to the consistency in U.S. performance on international tests since the 1960s (see **Ravitch, 2013**). Exaggerated claims about a lagging U.S. system often draw on PISA results (see, e.g., **Feuer, 2012**). It is also significant to note that U.S. has ranked relatively better on the Trends in Mathematics and Science Study (TIMSS), which assesses 4th and 8th grade in math and science, and Progress on International Reading Literacy Study (PIRLS), which assesses reading achievement of 4th graders, than on PISA. Not only is it important to look at different ILSA results, but continued discussion is also needed about what ILSAs do and do not measure (see, e.g., Kane, 2013 in **Chatterji, 2013**). For example, in Heckman and Kautz's (2013) **recent report**, they argue that achievement tests are unable to fully encapsulate valuable skills such as curiosity, motivation, and creativity (see also Perlman Robinson and Alexander's **discussion** of the importance of non-cognitive skills).

Myth #3: *ILSA results are predictive of long-term macroeconomic outcomes.* Based on this assumption, there is a projected image of U.S. educational decline and its negative economic implications. The now familiar alarmist rhetoric linking stagnating scores with a prediction of declining economic productivity is based on an assumed causal connection between PISA scores and long-term macroeconomic outcomes. Some researchers have called for greater caution in making such predictive and causal links, suggesting that "the discourse seems to run ahead of the evidence" (Feuer, 2013, p. 205 in **Chatterji, 2013**; see also **Kagan, 2012**).

Myth #4: *International benchmarking based on ILSA results provides sufficient evidence to enable the transfer of well-informed, best practices to American education.* One of the more significant ironies is that while U.S. education policy-makers frequently call for borrowing "best practices" from top-performing countries, U.S. educational reforms that emphasize high stakes testing as the principal tool of accountability represent the opposite of what top-performing countries actually do (see **Engel, Williams & Feuer, 2012**). International benchmarking, often based on average scores and league tables, is utilized as superficial "wake-up calls" to inspire system reform. This practice can undercut the potential of the information contained in the ILSAs to stimulate and facilitate deeper and more effective comparative investigations.

Myth #5: *Because of sampling or other methodological imperfections, ILSAs offer little or no value.* Fervent critiques of ILSAs tend to overstate their limitations and obscure the more subtle inferences that can be derived from rigorous comparisons. Comparing the large and fragmented

system in the U.S. with small and relatively homogeneous systems like Finland or Korea can be obviously fraught with complexity. But there is no question that with appropriate cautions there is much that can be learned from well-designed and executed cross-national assessments of student achievement.

Beyond the myths: *As is true for much of educational research and rhetoric, extreme positions limit the possibilities for evidence-informed progress. Sweeping claims that distort the evidence of achievement of U.S. students relative to other systems promotes either a kind of "sky is falling" rhetoric or becomes an invitation to apologize for an untenable status quo; and meanwhile the inherent value of rigorous comparisons is diluted or lost. It would be a mistake, though, to dismiss comparative research on the grounds that it doesn't enable singular and definitive conclusions. We believe the contrary is true: With less defensively held positions and greater balance, cross-national comparisons of student achievement offer exciting potential to educational research, policy, and practice.*

Laura C. Engel, The George Washington University

Michael J. Feuer, The George Washington University

Categories: [Validity](#) [High-stakes testing](#) [School accountability](#) [International assessments](#) [Test use](#)

Tags: [international large-scale assessments](#) [PISA](#) [TIMSS](#) [validity](#)

